DOCUMENT RESUME

ED 040 578 88 EM 008 004

AUTHOR Cook, J. Marvin
TITLE An Experimental Research Design for Comparing the Retention Effects of Different Instructional Treatments. An Occasional Paper.
INSTITUTION Maryland Univ., Baltimore.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
PUB DATE Feb 70
CONTRACT OEC-3-9-090018-0021(010)
NOTE 28p.

EDRS PRICE EDRS Price MF-$0.25 HC-$1.50
DESCRIPTORS *Comparative Analysis, Educational Research, *Post Testing, *Retention, Statistical Analysis, Teaching Methods

ABSTRACT

The effects of instructional treatments are most often evaluated on the basis of an analysis of posttests given immediately following the completion of the instruction. In order to compare the relative effectiveness of different methods of instruction, it is also useful to know something about the retention effects of various instructional treatments. Interpretation of the effect different teaching methods have on retention must be based, not merely on performance on the delayed posttest, but on the relationship of the performances on the immediate and the delayed posttests. An example of research involving four instructional treatments and two repeated measures (posttests) with no pretests is discussed as an application of repeated measures analysis to the study of retention. The equations necessary for determining certain critical factors in the analysis are demonstrated. The location of significant differences, calculatio n of the slopes of retention curves, and tests for assumptions of homogeneity of variance are also illustrated. A reference list is appended. (JY)

PRINTED BY THE STANDARD REGISTER COMPANY, U.S.A.
FORM 0801

ED040578

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

AN EXPERIMENTAL RESEARCH DESIGN FOR COMPARING THE RETENTION EFFECTS OF DIFFERENT INSTRUCTIONAL TREATMENTS

J. Marvin Cook

Associate Professor

An Occasional Paper

UNIVERSITY OF MARYLAND BALTIMORE COUNTY

DIVISION OF EDUCATION

Baltimore, Maryland

February, 1970

EM 008 004

ACKNOWLEDGEMENTS

Dr. Kenneth H. Wodtke of the Pennsylvania State University published an article entitled "On the Assessment of Retention Effects in Educational Experiments" in The Journal of Experimental Education, Volume 35, Number 4, Summer, 1967.

This paper is an interpretation of the ideas expressed by Dr. Wodtke.

The author wishes to acknowledge the significant contributions Dr. Wodtke's article has made to the development of the concepts in this paper.

The author also wishes to acknowledge and express appreciation for the theoretical assistance Dr. James Raths, Director of the Bureau of Educational Research and Field Services, University of Maryland, College Park, provided the author. Dr. Raths gave enthusiastic support throughout the writing of this paper.

The preparation of this paper was supported in part by U.S. Small Project Research Contract, Project No. 9-C-018, Contract No. OEC-3-9-090018-0021(010) from Regional Office III, Bureau of Research, U.S. Office of Education, U.S. Department of Health, Education and Welfare.

Comparing the Retention Effects of Different Instructional Treatments

The effects of instructional treatments are most often evaluated on the basis of an analysis of posttests given immediately following the completion of the instruction. In such analyses a comparison of treatment means is made to determine the relative value of the treatments in facilitating the amount of learning or the speed of learning. When the effect on learning is the point of interest, elementary ANOVA research designs may be utilized to compare the relative effectiveness of different methods of instruction. However, when there is an interest in investigating the retention effects of instructional treatments, the problem of analyses becomes more complex.

This paper describes in Part I some interpretative problems involved in the investigation of such retention effects. In Part II the use of the repeated measures analysis for studying retention is demonstrated. For the purpose of this paper the following definitions will be utilized:

Subject matter is defined in a broad sense as whatever it is you, the teacher, are intending to achieve in the way of educational intents. The educational intents might include among others, any combination of: knowledge, problem solving skill, creativity, or master skills.

Retention is defined in terms of the amount of learned subject matter that is retained after a specified interval of time.

Rate of forgetting is defined in terms of the change in the scores made on the immediate posttest and on the delayed posttest over a specified interval of time.

Over-all performance produced by a treatment is defined in terms of the score obtained when the scores made on the two or more posttests by the group of students under a particular treatment are pooled.

Part I

In terms of learning and retention there are several possible effects that might be produced by an instructional treatment. Although an instructional modification might be of little value in increasing the amount learned or the speed of learning, it might better facilitate retention. That is, while an instructional treatment might produce relatively inefficient learning, it might produce greater resistance to forgetting than some other methods. The opposite may also occur: An instructional treatment might facilitate the amount learned but provide very little resistance to forgetting. While producing differences in over-all performance, treatments may or may not produce different rates of forgetting. As previously mentioned, the over-all performance produced by an instructional treatment will be dependent upon its effect on resistance to forgetting.

Educational investigators wishing to compare the effects of different instructional treatments on learning and retention commonly use a modified form of the pre-posttest research design. Students are randomly assigned to two or more treatments. Depending upon whether or not the subject matter to be taught is familiar to the students, a pre-test may or may not be administered to them before beginning the experimental treatment. After the experimental treatments are administered, achievement tests (henceforth referred to as posttests in this paper) are then given to measure the performance level of the students. If the investigator is only interested in the effects produced by the treatments on learning, then one posttest may be given immediately after the instructional treatment. If the investigator is concerned with retention over time, then the immediate posttest may be administered plus one or more additional delayed

posttests administered at successive time intervals.

A discussion of three possible outcomes from a comparison of two instructional treatments will delineate the factors which must be considered when interpreting the comparative effects of treatments on retention. For simplicity and clarity this discussion will be limited to a two-treatment, two-posttest situation where the two posttests are separated by an interval of time. Figures 1, 2, and 3 contain retention curves (i.e., forgetting curves) obtained by plotting the data from the two successive post-test measures. These curves provide a visual indication of the rate of change in scores from the immediate posttest to the delayed posttest.

Outcome 1

Figure 1 is a plot of the tabulated data shown in Table 1. The data tabulated are the mean performance scores of the groups of students subjected to the treatments. There is statistically no difference between the effects of T_1 and T_2 on performance on the immediate posttest. However, T_1 produced a lower performance on the delayed posttest than T_2. Hence, the students subjected to T_1 reflect a higher rate of forgetting than the students subjected to T_2.

> Summarizing, there is no significant difference between the effect of the two treatments on learning; however, T_2 facilitates retention more than T_1.

There is a decrease of 12.5 from the mean score of immediate posttests to the mean scores of the delayed posttests (from 40.5 to 28). This pattern of loss does not hold for each of the two treatments. For T_1, the mean score decreases 20 from the immediate to the delayed posttests (from 40 to 20). For T_2, the score decreases only 5 (from 41 to 36). Since the over-all difference between performance on the immediate posttest and performance on the delayed posttest is not identically reflected by each treatment, there is an interaction between the treatments and the measures of retention

Table 1

Treatments	Mean Performance: Immediate	Mean Performance: Delayed	Over-all Performance Score
T_1	40	20	30
T_2	41	36	38.5
Means of posttests	40.5	28	

over time. The difference in the slopes of the retention curves in Figure 1 reflects the existence of the interaction. Figure 1 also reveals by the relative slope of the two curves that treatment T_2 produces greater resistance to forgetting than T_1.

Note that T_1 and T_2 not only produced differences in over-all performance, but also produced different rates of forgetting.

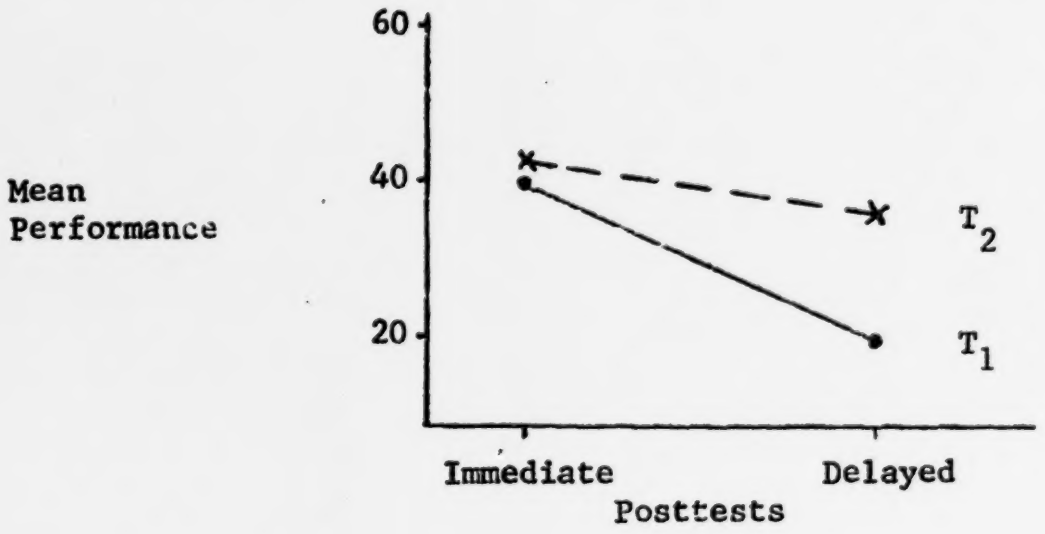

Figure 1

Outcome 2

Posttests of learning and retention are usually highly correlated because they must measure performance on the same subject matter. The problem of interpreting the scores on delayed retention posttests is a result of this fact. If two treatments yield differences in scores on the immediate posttest (reflecting a difference in amount learned), it could be expected because of the posttest correlation that the differences in scores would still exist on the delayed retention posttest. Misleading conclusions could be obtained if this high correlation between the learning and the retention posttests were ignored. The data from possible Outcome 2 illustrates the interpretive problem.

The tabulated data in Table 2 is plotted in Figure 2. There is a significant difference between the performance on the posttest of those students subjected to T_1 and the performance on the same test of the students subjected to T_2. This difference in performance is retained on the delayed posttest. Although the students who were taught under T_1 tested higher on both the immediate posttest and the delayed posttest, the fact that the forgetting curves of T_1 and T_2 are parallel (i.e., slopes are the same) indicate that the treatments did not produce different rates of forgetting. Hence, the existent differences on the delayed posttest is attributable to the differing effect of the treatments on the amount learned and not to a differing effect on retention.

> Summarizing, T_1 facilitated learning more than T_2 but apparently had no advantage over T_2 in terms of retention.

Table 2

Treatments	Mean Performance		Over-all Performance Score
	Immediate	Delayed	
T_1	60	50	55
T_2	45	35	40
Means of posttests	52.5	42.5	

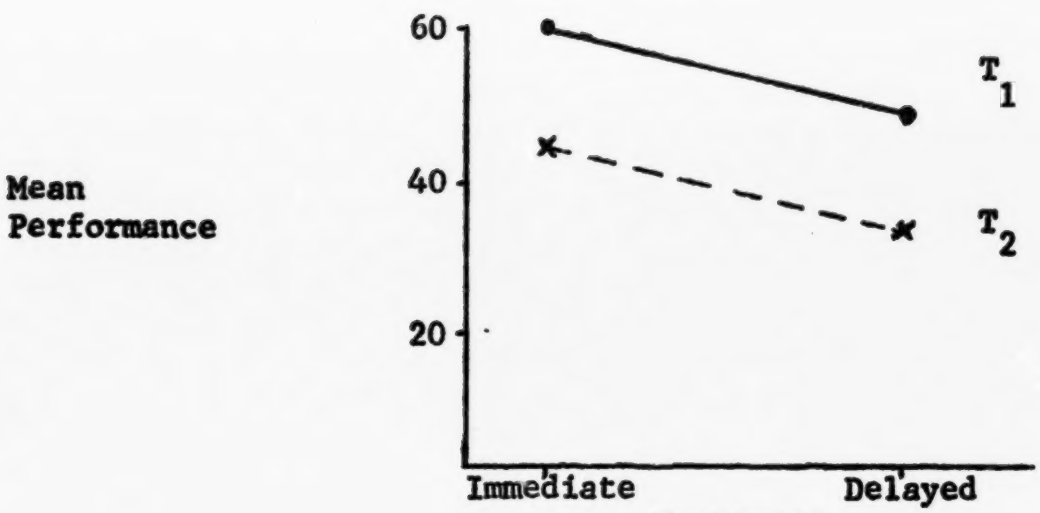

Figure 2

Note that the decrease of 10 from the mean score on the immediate posttests to the mean score of the delayed posttests (i.e., from 52.5 to 42.5) is identically reflected by each treatment. Hence, there is no interaction between the treatments and the measures of retention over time. The parallel retention curves in Figure 2 reflect this condition.

It is also worthwhile to note that while T_1 and T_2 produce differences in over-all performance, the treatments did not produce different rates of forgetting.

Outcome 3

An interesting and important question for instructional modification

is whether an instructional treatment facilitates retention independent of its effect on learning. The data from possible Outcome 3 lends itself to a consideration of this question.

The data of a third possible outcome is tabulated in Table 3 and plotted in Figure 3. The data reveals a difference between the performance of students under T_1 and the performance of students under T_2 on the immediate posttest. However, a difference is essentially non-existent on the delayed posttest. The data shows that the treatments produced differences in learning, but that the differences were cancelled over time. Hence, the apparent advantage that T_2 illustrates over T_1 in learning is wiped out when the two treatments are compared on the basis of the delayed posttest.

But as we have seen in our discussion of the previous two possible outcomes, it is not sufficient to evaluate treatments just on the basis of the scores made on delayed posttests.

> The relation of these two sets of scores must be first determined for a proper evaluation to be made.

Note that the decrease of 15.5 from the mean score of the immediate posttests to the mean score of the delayed posttests(i.e., from 55 to 39.5) is not identically reflected on each of the two treatments. Therefore, there is an interaction between the treatments and the measures of retention over time. Figure 3 reflects this interaction by showing the differences in slopes of retention curves. In this outcome T_1 produced a relatively lower rate of forgetting than T_2.

Table 3

Treatments	Mean Performance: Immediate	Mean Performance: Delayed	Over-all Performance Score
T_1	45	39	42
T_2	65	41	53
Means of posttests	55	39.5	

As in Outcome 1, the data in Outcome 3 reveals that a difference existed not only between the over-all performances produced by T_1 and T_2 but also between the rates of forgetting produced by T_1 and T_2.

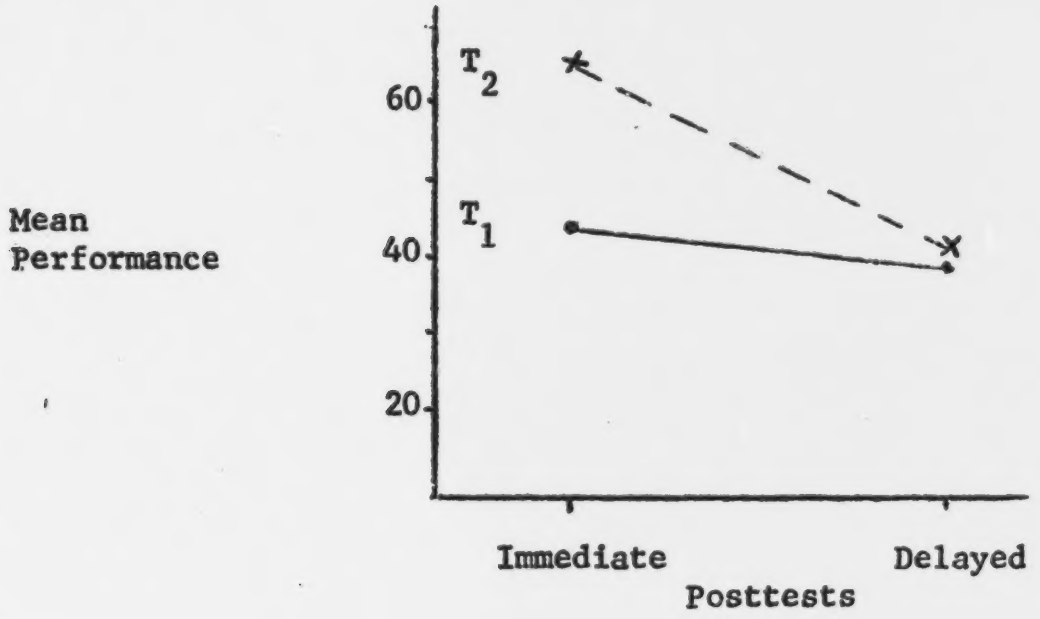

Figure 3

Conclusion

These three possible effects illustrate the point that interpretation of the effect different teaching methods have on retention must be based not merely on performance on the delayed posttest (i.e., the retention measure) but on the relationship of the performances on the immediate and the delayed posttests.

> It is the rate of forgetting that is the key criteria for determining whether or not a particular treatment facilitates retention more than another treatment.

The existence of a statistically significant treatment-by-retention measure interaction indicates that there is a difference in the rate of forgetting produced by the treatments. A plot of the retention curves reveals which treatment produces the lower rate of forgetting.

Part II

The thrust of the discussion in Part I was that the analysis of retention data is basically a problem of comparing the rates of forgetting curves. Wodtke (1) states that the repeated measures design seems most appropriate for the study of differential rates of forgetting. Excellent descriptions of repeated measures design have been given by Grant (2) and Winer(3).A repeated measures analysis provides evidence concerning the general superiority in terms of over-all performance between treatments and it also provides tests for the differences in the slopes of the forgetting curves. An analysis of the general superiority between treatments is based upon statistical differences between the over-all performance produced by each instructional treatment.

For the purposes of demonstrating the application of the repeated measures analysis to studying retention, an example research involving four instructional treatments and two repeated measures (posttests) with no pre-test will be discussed. Note that only one delayed posttest will be utilized. If more than one delayed posttest were to be administered, a trend analysis could be performed to determine the shape of the forgetting curves. A discussion of such an analysis is given by Grant (2). Winer (3) describes a repeated measures analysis design appropriate for a study which includes a pre-test and two or more posttests.

Example

A researcher wished to test the following two research hypotheses:

1) Four instructional treatments A1, A2, A3, and A4 have differential effects on the over-all performance of students.
2) The four treatments produce different rates of forgetting.

He classified the 72 students in the fourth-grade classes of the ABC Elementary School into three intelligence levels - high, middle, and low. The students who scored above 120 on the Wisc , between 100 and 120 on the Wisc, and below 100 on the Wisc were respectively identified as high, middle and low. The students from each of the ability levels were then randomly assigned to the three treatments. While the n's in each sub-level were not equal, they were not inordinately low. The layout for this study appears as follows:

		Posttests	
		Immediate	Delayed
Treatment A1	H1	X	
	M1		
	L1		
Treatment A2	H2		
	M2		
	L2		
Treatment A3	H3		
	M3		
	L3		
Treatment A4	H4		
	M4		
	L4		

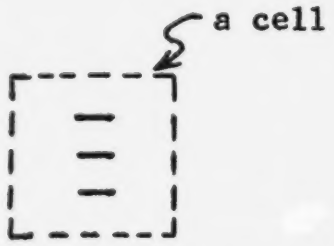

Figure 4

The X shown in the layout is an example of datum entered into a particular location in the layout. X located as shown would be the mean score of the students of high ability level (H1) under treatment A1 on the immediate posttest.

In this design, an individual's score is assumed to be represented as follows:

$$X_{jkp} = \mu + \alpha_k + \beta_j + \phi_{k(p)} + \gamma_{jk} + \beta\phi_{jk(p)} + \varepsilon_{jkp}$$

where:

X_{jkp} is the score found under the j^{th} posttest; in the k^{th} treatment group; and in the l^{th} ability level.

μ is the population mean.

α_k is the effect of being in the k^{th} treatment group. This effect in this layout is estimated by

$$\hat{\alpha}_{A1} = \bar{X}_{A1} - \bar{X}_T \qquad \hat{\alpha}_{A3} = \bar{X}_{A3} - \bar{X}_T$$
$$\hat{\alpha}_{A2} = \bar{X}_{A2} - \bar{X}_T \qquad \hat{\alpha}_{A4} = \bar{X}_{A4} - \bar{X}_T$$

β_j is the effect of the performance measure being posttest j. It is estimated in this example by subtracting each posttest mean from the grand mean:

$$\hat{\beta}_{Immed.} = \bar{X}_{Immed.} - \bar{X}_T \quad \text{and} \quad \hat{\beta}_{Del.} = \bar{X}_{Del.} - \bar{X}_T$$

$\phi_{k(p)}$ is the effect of being in the p^{th} level of ability within treatment k. This effect is estimated by subtracting each level mean within a treatment from its treatment mean.

γ_{jk} is the interaction of treatments and repeated performance measures (posttests) over time. It is estimated in this example layout by finding the mean score of all the scores in a cell consisting of a specific treatment group and a particular posttest. The dotted lines in the layout define such a cell. The cell effect is estimated by subtracting the grand mean $\bar{\bar{X}}_T$ from this mean score. The alpha effect α_k and the beta β_j for this cell are also subtracted out and the resulting number is the estimated interaction effect $\hat{\gamma}_{jk}$.

$$\text{i.e. } \hat{\gamma}_{jk} = (\bar{x}_{jk} - \bar{x}_T) - \alpha_k - \beta_j$$
$$= (\bar{x}_{jk} - \bar{x}_T) - (\bar{x}_k - \bar{x}_T) - (\bar{x}_j - \bar{x}_T)$$
$$= \bar{x}_{jk} - \bar{x}_k - \bar{x}_j + \bar{x}_T$$

$\beta\phi_{jk(p)}$ is the interaction of levels of ability and performance on posttests within a particular treatment. The interaction effect can be estimated for each datum in the analysis by the following formula:

$$\beta\phi_{jk(p)} = \bar{x}_{jkp} - \bar{x}_p - \bar{x}_{jk} + \bar{x}_k$$

ε_{jkp} is the error term.

The <u>assumptions</u> made for the repeated measure analysis were:

1. The students from each of the ability levels were randomly assigned to the four treatments.
2. The variances of the ability level means within the various treatments are homogeneous.
3. The variance of the "interaction of posttests and ability levels within treatments" is a pooling of the variance of the scores in each treatment group about the treatment mean after the effects of posttests and the effects of ability levels within treatments have been subtracted. These variances for each treatment group are homogeneous.
4. The data collected satisfies at least the interval scale.
5. The scores collected as data are normally distributed.

The research was designed to test the following two <u>statistical hypotheses:</u>

For row effects:

Null: $\alpha_k = 0$, for all k

Alternate: $\alpha_k \neq 0$, for some k

For interaction effects:

Null: $\gamma_{jk} = 0$, for all jk

Alternate: $\gamma_{jk} \neq 0$, for some jk

Findings and Analysis:

Treatments	Levels	Posttests: Immediate	Posttests: Delayed	Total	
A1	H1	45	21	66	
	M1	40	15	55	150
	L1	25	4	29	
A2	H2	87	78	165	
	M2	77	67	144	413
	L2	58	46	104	
A3	H3	49	19	68	
	M3	46	27	73	182
	L3	30	11	41	
A4	H4	96	93	189	
	M4	88	83	171	494
	L4	70	64	134	
		$\Sigma X_I = 711$	$\Sigma X_D = 528$	1239	

1. Total sum of squares (SST)

$$SST = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

$$= 82,605 - \frac{(1239)^2}{24} = 82,605 - 63,735$$

$$= 18,870$$

2. Sum of squares between ability levels within all treatments (the variability between ability level means and the grand mean).

$$SS_{Between\ levels} = \frac{(\Sigma X_{H1})^2}{n_{H1}} + \frac{(\Sigma X_{M1})^2}{n_{M1}} + \ldots + \frac{(\Sigma X_{L4})^2}{n_{L4}} - \frac{(\Sigma X)^2}{N}$$

$$= \frac{(66)^2}{2} + \frac{(55)^2}{2} + \ldots + \frac{(134)^2}{2} - 63,735$$

$$= 17,040.5$$

3. Sum of squares between treatments (the variability between treatment means and the grand mean).

$$SS_{\text{Between treatments}} = \frac{(\Sigma X_{A1})^2}{n_{A1}} + \ldots + \frac{(\Sigma X_{A4})^2}{n_{A4}} - \frac{(\Sigma X)^2}{N}$$

$$= \frac{(150)^2}{6} + \ldots + \frac{(494)^2}{6} - 63{,}735$$

$$= 14{,}636.5$$

4. Sum of squares for levels of ability within treatments (the variability of ability level means from treatment means).

$$SS_{\text{Levels within treatments}} = SS_{\text{Between ability levels}} - SS_{\text{Between treatments}}$$

$$= 17{,}040.5 - 14{,}636.5$$

$$= 2{,}404.0$$

5. Sum of squares within ability levels (variability of ability level scores from ability level mean).

$$SS_{\text{Within levels}} = SST - SS_{\text{Between ability levels}}$$

$$= 18{,}870.0 - 17{,}040.5$$

$$= 1{,}829.5$$

6. Sum of squares for posttests (variability of posttest means from the grand mean).

$$SSP = \frac{(\Sigma X_I)^2}{n_I} + \frac{(\Sigma X_D)^2}{n_D} - \frac{(\Sigma X)^2}{N}$$

$$= \frac{(711)^2}{12} + \frac{(528)^2}{12} - 63{,}735$$

$$= 1{,}623.8$$

7. Sum of squares for cells (variability of cell means from the grand mean). The quantity is not used in the summary table; it will be used to find interaction sum of squares.

$$SS_{\text{Cells}} = \frac{(\Sigma X_{A1I})^2}{n_{A1I}} + \frac{(\Sigma X_{A1D})^2}{n_{A1D}} + \frac{(\Sigma X_{A2I})^2}{n_{A2I}} + \ldots + \frac{(\Sigma X_{A4D})^2}{n_{A4D}} - \frac{(\Sigma X)^2}{N}$$

$$= \frac{(110)^2}{3} + \frac{(40)^2}{3} + \frac{(222)^2}{3} + \ldots + \frac{(240)^2}{3} - 63{,}735$$

$$= 16{,}416.7$$

8. Sum of squares for the treatment and posttests interaction (variability left in a cell after the treatment effects and the posttests effects have been subtracted out).

SSInteraction = SSCells - SSP - SSBetween treatments

= 16,416.7 - 1,623.8 - 14,636.5

= 156.4

9. Sum of squares within treatment groups (the sum of the variabilities of scores within each treatment group from the treatment group mean).

$$SS\text{Within treatment groups} = \Sigma X^2 - \frac{(\Sigma X_{A1})^2}{n_{A1}} - \ldots - \frac{(\Sigma X_{A4})^2}{n_{A4}}$$

$$= 82{,}605 - \frac{(150)^2}{6} - \ldots - \frac{(492)^2}{6}$$

= 4,233.5

10. Sum of squares for the interaction between posttests and the effects of levels of ability within treatments (the variability within each treatment group after the effects of posttests and the effects of levels within treatments have been subtracted out).

SSInteraction of posttests and levels within treatments =

SSWithin treatment groups - SSP - SSLevels within treatments - SSInteraction.

= 4,233.5 - 1,623.8 - 2,404.0 - 156.4

= 49.3

Table 4

Summary Table Model (for k treatments, j posttests, and p=km total levels where m is the number of levels within each treatment)

	Source	df	
	Between Levels	p-1	
	Between treatments	k-1	-- (use Error between to test this MS)
(Error between)	Levels within treatments	p-k	
	Within Levels	p(j-1)	
	Posttests	j-1	-- (Use Error within to test this MS)
	Interaction: (Treatment x post tests)	(k-1)(j-1)	-- (Use Error within to test this MS)
(Error within)	Interaction: (Posttests x Levels within treatments)	(j-1)(p-k)	
	Total	N-1	

TABLE V

SUMMARY TABLE FOR THIS EXAMPLE

Source	df	SS	MS	F	F @ .05
Between Levels	11	17,040.5			
Between treatments	3	14,636.5	4,878.8	16.24	4.06
Levels within treatments [error (between)]	8	2,404.0	300.5		
Within Levels	12	1,829.5			
Posttests	1	1,623.8	1,623.8	-	
Treatment x Posttests	3	156.4	52.1	8.54	4.06
Posttests x Levels w. Treatment [error (within)]	8	49.3	6.1		
	23	18,870.0			

Findings: (1) The null hypothesis $\alpha_k = 0$ is rejected.

(2) The null hypothesis $\gamma_{jk} = 0$ is rejected.

Conclusions: (1) The research hypothesis that the four instructional treatments have differential effects on the over-all performance of the students is supported by the data of this experiment.

(2) The research hypothesis that the four treatments result in different rates of forgetting is supported by these data.

Related Findings:

1. Location of significant differences.

By location the significant differences between means, the researcher can determine which of the four treatments had the greater effect on the over-all performance of the students. The following conditions hold in this experiment:

i) the maximum probability of making a type-I error (α) is set at .05,

ii) the planned contrasts are not orthogonal,

iii) the error rate (α) has an experiment base,

iv) contrasts with other than the control group are to be made,

v) the number of comparisons c is 6,

vi) the number of treatments k is 4,

vii) the number of comparisons, c = 6, is equal to $k(k-1)/2$, and

viii) only pair-wise contrasts are to be made.

Based upon the above conditions and the schema recommended by Hopkins and Chadbourn (4) for making c multiple comparisons among k treatment means, the Tukey (b) method described by Winer (5) may be applied to locate significant differences between means.

The Tukey (b) method uses a studentized range statistic defined by

$$q_{obs} = \frac{\bar{X}_1 - \bar{X}_s}{\sqrt{MSE/n}}$$

where n is the number of observations in each $\bar{X}$. The studentized range statistic is the quotient obtained by dividing the difference between the largest and smallest treatment means (the range of the treatment means) by the within group variance (often referred to as the error mean square, MSE) of the ANOVA procedure over n.

The Tukey (b) test is a compromise between the Newman-Keuls (6) procedure and the more conservative Tukey (a) procedure (i.e., more conservative in terms of keeping the type-I error small for a given alpha). The q-critical of the Tukey (b) method is given by

$$q_c = \frac{q_{1-\alpha}(k, f) + q_{1-\alpha}(r, f)}{2}$$

where k = number of treatments (i.e., the range of the number of steps between means), f = degrees of freedom for MSE, and r = number of steps between two means on an ordered scale which are being compared. The number of steps r between two means X_j and X_i is j - i + 1 where j and i are the rank order of the means being compared. The symbol $q_{.99}(k, f)$ designates the 99th percentile point on the q distribution. Tables of the studentized range statistic q distribution containing critical values of $q_{1-\alpha}(k, f)$ are given in Table B.4 in Winer (3). Critical values of $q_{1-\alpha}(r, f)$ are obtained from the same tables by setting r equal to the range.

If the observed studentized range statistic q_{obs} is larger than q_c, the difference between the means is significant.

The following calculations demonstrate a comparison between the means of treatment A1 and A2:

To find q_{obs}:

$$\bar{X}_{A1} = \frac{150}{6} = 25.0$$

$$\bar{X}_{A2} = \frac{413}{6} = 68.8$$

$$n = 6$$

$$MSE = 300.5$$

$$q_{obs} = \frac{68.8 - 25.0}{\sqrt{300.5/6}} = 6.19$$

To find q_c:

Placing the mean scores in rank order we have,

$\bar{X}_{A1}$	$\bar{X}_{A3}$	$\bar{X}_{A2}$	$\bar{X}_{A4}$
25.0	33.3	68.8	82.3

Hence $k = 4$

$r = 3 - 1 + 1 = 3$

$f = 8$

with $q_{.95}\ (4, 8) = 4.53$

and $q_{.95}\ (3, 8) = 4.04$

Thus $q_c = \frac{4.53 + 4.04}{2} = 4.29$

Since the observed q statistic exceeds the critical value, the difference between the means is significant. Pair-wise comparisons of other treatments are made in a similar manner.

2. Slopes of Retention Curves.

The fact that there is a statistically significant treatment x repeated performance measure interaction indicates that the slopes of the retention curves for the treatment groups are different. To determine which treatments result in decreased rate of forgetting (i.e., improve retention), the slopes of the retention curves may be plotted as shown in Figure 5 based upon the data in Table VI.

TABLE VI

	Immediate	Delayed
Treatment A1	37	13
Treatment A2	75	64
Treatment A3	42	19
Treatment A4	85	80

Note: The values used in Table VI are means calculated from the data shown on page 13.

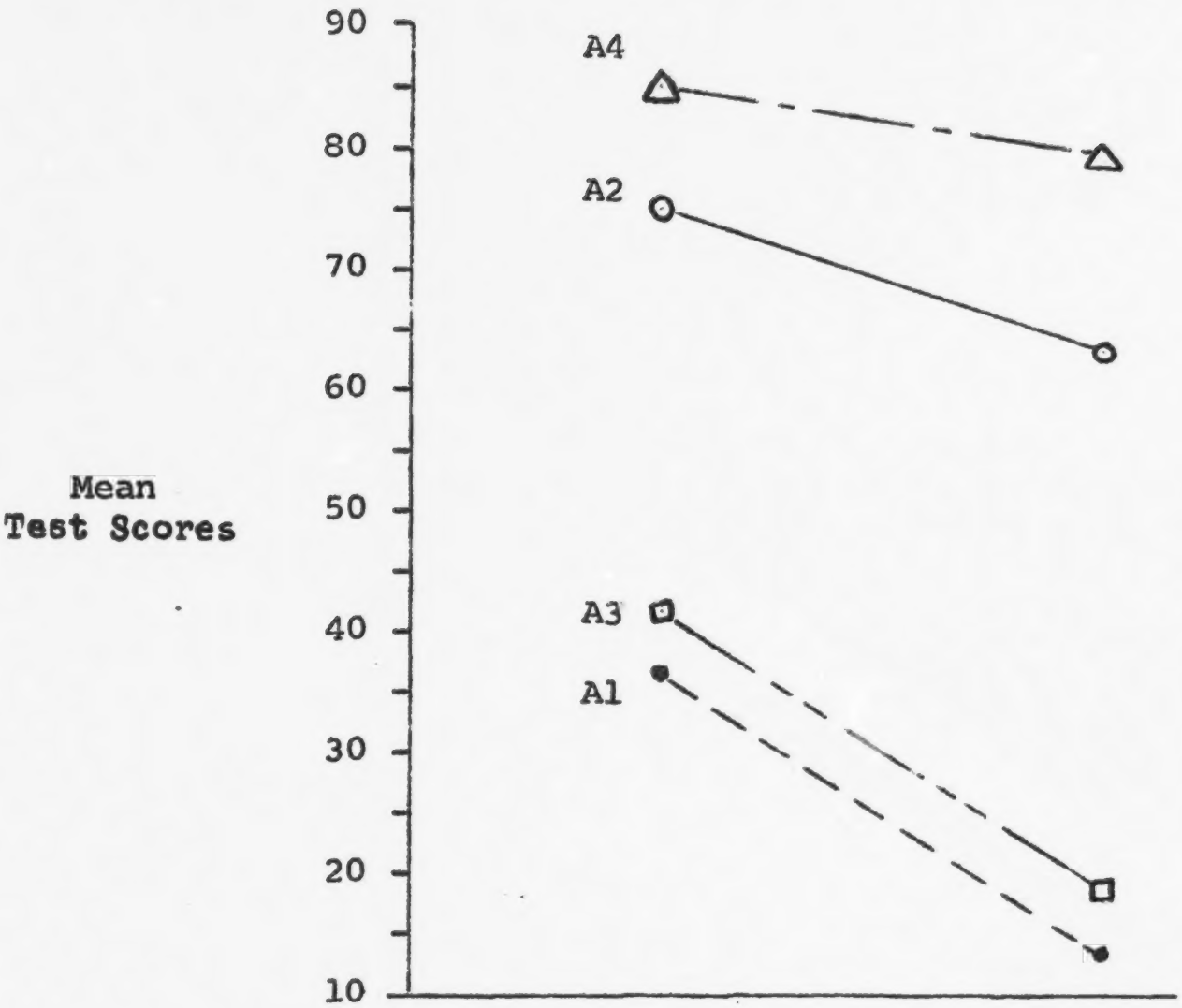

Figure 5

The slopes of the retention curves of treatments A1 and A3 are essentially the same. Therefore, treatment A3 offers no advantage over treatment A1 in terms of improvement of retention. However, the slopes of A2 and A4 are not as steep as A1 and A3 and, hence, represent an improvement in retention.

The slope of A4 being less than the slope of A2 says that the rate of forgetting of A4 is less than that

of A2 . That is, A4 results in greater improvement of retention than any of the other three treatments.

3. Estimation of effects:

$$\text{Effect of A1} = \hat{\alpha}_{A1} = \bar{X}_{A1} - \bar{X}_T = 25.0 - \frac{1239}{24} = -26.6$$

$$\text{Effect of A2} = \hat{\alpha}_{A2} = 68.8 - 51.6 = 17.2$$

$$\text{Effect of A3} = \hat{\alpha}_{A3} = 33.3 - 51.6 = -18.3$$

$$\text{Effect of A4} = \hat{\alpha}_{A4} = 82.3 - 51.6 = 31.7$$

Interaction effect associated with treatment A1 and immediate posttest:

$$\gamma_{A1\text{-}I} = \bar{X}_{A1\text{-}I} - \bar{X}_I - \bar{X}_{A1} + \bar{X}_T$$

$$= 36.7 - 59.3 - 25.0 + 51.6$$

$$= 4.0$$

4. Tests for assumptions of homogeneity of variance:

(a) The variance "levels within treatments" is a pooling of the variances of the level means from the treatment mean within treatment A1 ; plus the variance of the level means from the treatment means within treatment A2 , etc. The variances of the level means within the various treatments must be homogeneous to have a valid test of the treatment effects. This assumption is tested by finding

the variances of the level means about each treatment mean and testing the homogeneity of these variances by an F-test. For this example research:

$$\text{SS of levels within A1} = \frac{(\Sigma X_{A1\text{-}H1})^2}{n_{A1\text{-}H1}} + \frac{(\Sigma X_{A1\text{-}M1})^2}{n_{A1\text{-}M1}} + \frac{(\Sigma X_{A1\text{-}L1})^2}{n_{A1\text{-}L1}} - \frac{(\Sigma X_{A1})^2}{n_{A1}}$$

$$= \frac{(66)^2}{2} + \frac{(55)^2}{2} + \frac{(29)^2}{2} - \frac{(150)^2}{6} = 361$$

$$\text{SS of levels within A2} = \frac{(165)^2}{2} + \frac{(144)^2}{2} + \frac{(104)^2}{2} - \frac{(413)^2}{6} = 960.3$$

$$\text{SS of levels within A3} = \frac{(68)^2}{2} + \frac{(73)^2}{2} + \frac{(41)^2}{2} - \frac{(182)^2}{6} = 296.3$$

$$\text{SS of levels within A4} = \frac{(189)^2}{2} + \frac{(171)^2}{2} + \frac{(134)^2}{2} - \frac{(494)^2}{6} = 786.3$$

$$\text{MS of levels within A1} = \frac{361}{df} = \frac{361}{2} = 180.5$$

$$\text{MS of levels within A2} = \frac{960.3}{2} = 480.1$$

$$\text{MS of levels within A3} = \frac{296.3}{2} = 148.1$$

$$\text{MS of levels within A4} = \frac{786.3}{2} = 393.1$$

$$F_{max} = \frac{480.1}{148.1} = 3.24$$

The significance of this ratio is tested by entering the F-table for the df of the numerator and the df of the denominator--in this case 2,2 df.

F_{max} critical with 2,2 df at .05 level = 19.0

The assumption of homogeneity of the level within treatment variances is supported by these data.

(b) The variance "interaction of posttests and levels within treatments" is a pooling of the variance of the scores in each treatment group about the treatment mean after the effects of posttests and the effects of levels within treatments have been subtracted out. These variances for each treatment group must be homogeneous to have valid tests of the posttests mean square and the treatment x posttests mean square. This assumption is tested as follows: For each treatment, the sum of squares for cells is determined, and the SS for posttests and the SS for levels within treatments is subtracted out to find the variance of the interaction of posttests and levels within treatments for that treatment. The variances found for each treatment are then compared for homogeneity by the F-test. For this example research:

For treatment A1:

$$\text{SS Cells} = (45)^2 + (40)^2 + (25)^2 + (21)^2 + \ldots - \frac{(150)^2}{6} = 1182$$

$$\text{SS Posttests} = \frac{(110)^2}{3} + \frac{(40)^2}{3} - \frac{(150)^2}{6} = 816.6$$

$$\text{SS Levels within A1} = \frac{(66)^2}{2} + \frac{(55)^2}{2} + \frac{(29)^2}{2} - \frac{(150)^2}{6} = 361$$

SS of interaction between posttests and levels within treatments for treatment A1 = 1182 - 816.6 - 361 = 4.4

The sum of squares of the "posttests by levels within treatments" interaction may be calculated for the other treatments in a similar manner.

A2 = 3.2

A3 = 41.4

A4 = 3.3

These sum of squares are converted into variances by dividing by the appropriate degrees of freedom--in this example the product of (number of posttests - 1) times (number of levels in a treatment - 1) = 2 .

Variance of A1 = 2.20

Variance of A2 = 1.60

Variance of A3 = 20.70

Variance of A4 = 1.65

$F_{max} = \frac{20.70}{1.60} = 12.94$

F_{max} critical at 2,2 df @ .05 = 19.

The assumption of homogeneity of the posttests x levels within treatments variances is supported by the data of this experiment.

REFERENCES

1. Wodtke, Kenneth H. "On the Assessment of Retention Effects in Educational Experiments," The Journal of Experimental Education, XXXV, No. 4 (Summer, 1967), 30.

2. Grant, D. A. "Analysis of Variance Tests in the Analysis and Comparison of Curves," Psychological Bulletin (1956), pp. 141-54.

3. Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill Book Company, 1962, pp. 298-378.

4. Hopkins, Kenneth D., and Chadbourn, Russell A. "A Schema for Proper Utilization of Multiple Comparisons in Research and a Case Study," American Educational Research Journal, IV, No. 4 (November, 1967), 407-12.

5. Winer, Statistical Principles in Experimental Design, p. 87.

6. Ibid., pp. 77-85